

1/10

Figure 1: Top blastp results for INSP161 polypeptide sequence (SEQ ID NO: 18) against the NCBI-nr database.

BLASTP 2.2.2 [Jan-08-2002]

Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.

Query= insp161.pep
(470 letters)

Database: All non-redundant GenBank CDS
translations+PDB+SwissProt+PIR+PRF
1,523,012 sequences; 490,363,361 total letters

Searching.....done

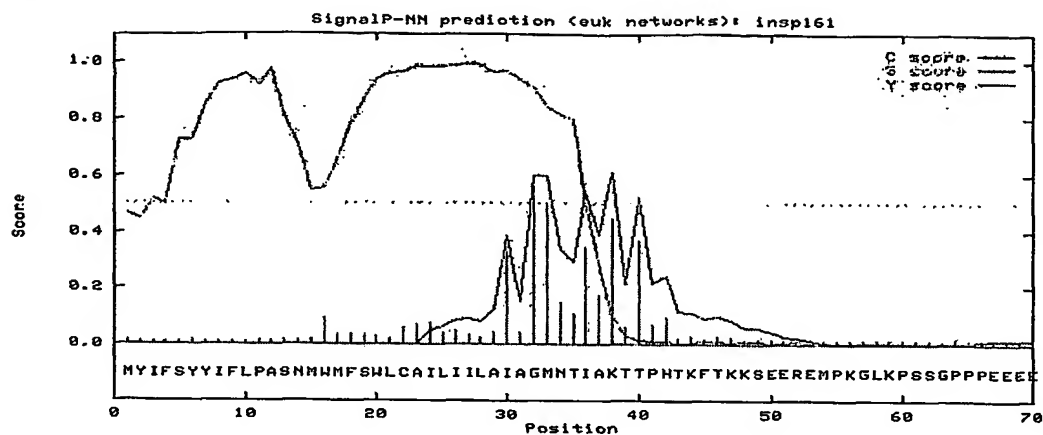
	Score	E
Sequences producing significant alignments:	(bits)	Value
ref XP_067228.5 similar to otolin-1 [Homo sapiens]	907	0.0
ref XP_143327.1 similar to INNER EAR-SPECIFIC COLLAGEN PRECURSO...	622	e-177
ref XP_227256.1 similar to INNER EAR-SPECIFIC COLLAGEN PRECURSO...	530	e-149
sp P83371 OTO1_ONCKE Otolin-1 precursor >gi 18496364 dbj BAB8456...	365	e-100
sp P98085 COLE_LEPMA INNER EAR-SPECIFIC COLLAGEN PRECURSOR (SACC...	351	2e-95
pir A55797 collagen precursor, saccule-specific - bluegill >gi ...	325	1e-87
ref XP_224253.1 similar to hypothetical protein FLJ31208 [Homo ...	257	4e-67
ref NP_848635.1 hypothetical protein MGC48915 [Homo sapiens] >g...	249	6e-65
ref XP_290602.1 similar to Adiponectin precursor (30 kDa adipoc...	249	1e-64
ref NP_777059.1 collagen, type X, alpha 1 (Schmid metaphyseal c...	246	6e-64

2/10

Figure 2: Signal peptide prediction (SignalP V2.0) output for INSP161 polypeptide sequence (SEQ ID NO: 18).

```
>insp161
```

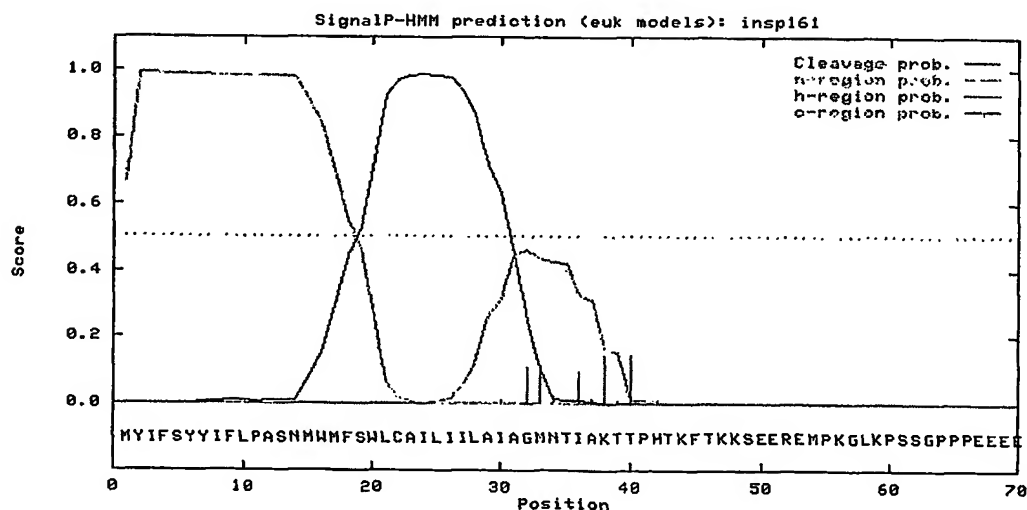
SignalP-NN result:



```
# data
```

```
>insp161
# Measure Position Value Cutoff signal peptide?
max. C 32 0.564 0.33 YES
max. Y 38 0.613 0.32 YES
max. S 28 0.992 0.82 YES
mean S 1-37 0.803 0.47 YES
# Most likely cleavage site between pos. 37 and 38: TIA-KT
```

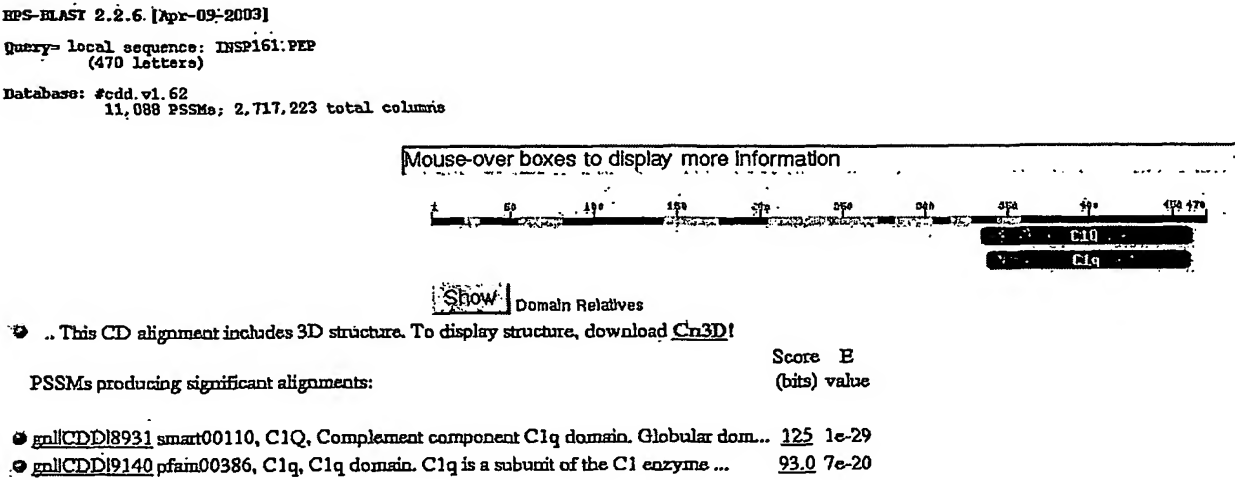
SignalP-HMM result:



```
# data
```

```
>insp161
Prediction: Signal peptide
Signal peptide probability: 0.667
Signal anchor probability: 0.326
Max cleavage site probability: 0.146 between pos. 37 and 38
```

Figure 3: NCBI CDD output for INSP161



4/10

Figure 4: INSP161 nucleotide and amino acid sequence

1	gcataatcatt cattcttatg attcatatac atatatattt	cttttatatg agtatatgtt
61	tatatgactt agtcttttca tttggacctc acataatatg	tatatatttt cctattatat
		m y i f s y y
	Signal peptide	
121	ctttcttcca gcttcaaata tgtggatggtt ttcttggctt	tgtgctattt taattatttt
	i f l p a s n m w m f s w l c a i l i i	
181	ggctattgct ggtatgaaca caatagcaaa	gaccacacca catacaaat ttacgaagaa
	l a i a g m n t i a	k t t p h t k f t k
241	atctgaggaa agagagatgc caaagggctt	aaagccatcc agtggcccac ctccagaaga
	k s e e r e m p k g l k p s s g p p p e	
301	agaagaaacc ctcttcacag aaatggctga	aatggcagaa ccaattacca aaccctcggc
	e e e t l f t e m a e m a e	p i t k p s
361	cttggattct gtctttggca ctgccactct	ctctcccttt gaaaacttca ctcttgaccc
	a l d s v f g t a t l s p f e n f t l d	
421	agctgatttc tttttgaatt gttgtgattg	ttgttcacct gtaccggggc agaaaggaga
	p a d f f l n c c d c c s p v p	g q k g
481	acctggagag actggacagc caggtcctaa	aggagaggct ggaaatttgg ggateccagg
	e p g e t g q p g p k g e a	g n l g i p
541	gccaccagga gttgttgggc cccaaggccc	tagaggctac aaaggagaga aagggtgaacc
	g p p g v v g p q g p r g y k g e k g e	
601	tggccctaag ggagataaag gaaacattgg	tttgggagga gtgaaaggac aaaaaggctc
	p g p k g d k g n i g l g g v k g q k g	
661	caaggagagac acatgtggga attgtacca	aggagaaaaa ggagaccaag gggctatggg
	s k g d t c g n c t k g e k g d q g a m	
721	ctcacctggc ctgcacggag ggcctggcgc	caagggagag aagggggaga tgggggagaa
	g s p g l h g g p g a k g e k g e m g e	
781	gggggagatg ggggataaag gctgctgtgg	agattctggg gagaggggag gaaaaggaca
	k g e m g d k g c c g d s g e r g g k g	
841	gaaagggtgag gggggtatga aaggggaaaa	aggtagcaaa ggagacagtg gaatggaagg
	q k g e g g m k g e k g s k g d s g m e	
901	caaaagcggc cgtaatggtc tgcctggggc	caaagggtgat ccagggatta aaggagaaaa
	g k s g r n g l p g a k g d p g i k g e	
961	aggagagtta ggtcctcctg gtctcctggg	acctactggg ccgaagggtg acattggcaa
	k g e l g p p g l l g p t g p k g d i g	
1021	caaaggggtc cgaggcccca ctgggaagaa	gggctctcgg ggctttaaag gctccaaggg
	n k g v r g p t g k k g s r g f k g s k	

Collagen domains

5/10

1081 tgagttggct agagtgcccc ggtcggcttt cagcgctggt ttgtcaaagc catttcctcc
g e l a r v p r s a f s a g l s k p f p

1141 tcctaacatc cccatcaaat ttgaaaagat tctctataat gaccaaggga attacagtcc
p p n i p i k f e k i l y n d q g n y s

1201 tgtcactggg aagtttaact gctctattcc tgggacatat gttttttcct accatattac
p v t g k f n c s i p g t y v f s y h i

1261 ggtgaggggg cgacctgctc gaatcagtct ggtggcccag aataagaagc agttcaagtc
t v r q r p a r i s l v a q n k k q f k

1321 cagagaaact ctctatggtc aggaaataga ccaggcctct ctcctcgtca tcttgaaatt
s r e t l y g q e i d q a s l l v i l k

1381 aagtgcagga gaccaagtct ggcttgaggt gtcaaaagat tggaatggg tgtatgtcag
l s a g d q v w l e v s k d w n g v y v

1441 tgctgaggat gacagcattt ttactgggtt cttttgtac ccagaggaaa cttctggaat
s a e d d s i f t g f l l y p e e t s g

1501 ttaccataa atttgtgtcc tgaatcctgt agtttagatt cagtggaata agtcagttaa
i s p

1561 cacagagtag tgctattaaa aaataacttc cattttttca agattat

C1q domain

Figure 5

A)

1 SSGPPPEEEE TLFTEMAEMA EPITKPSALD SVFGTATLSP FENFTLDPAD

51 FFLNCCDCCS PVF **GQKGEPG ETGQPGPKGE AGNLGIPGPP GVVGPQGPRG**
Collagen region

101 **YKGEKGEPGP KGDKGNIGLG GVKGQKSGK DTCGNCTKGE KGDQGAMGSP**
Collagen region

151 **GLHGGPGAAG EKGEMGEKGE MGDKGCCGDS GERGGKGQKG EGGMKGEKGS**
Collagen region

201 **KGDSGMEGKS GRNGLPGAAG DPGIKGEKGE LGPPGLLGPT GPKGDIGNKG**
Collagen region

251 **VRGPTGKKGS RGFKGSKGEI** ARVPR **SAFSA GLSKPFPPPN IPIKFEKILY**
C1q domain

301 **NDQGNYSPTV GKFNCSIPGT YVFSYHITVR GRPARISLVA ONKKQFKSRE**
C1q domain

351 **TLYGQEIDQA SLIVILKLSA GDQVWLEVSK DWNGVYVSAE DDSIFTGFLI**

401 YPEETSGISP

B)

1 SSGPPPEEEE TLFTEMAEMA EPITKPSALD SVFGTATLSP FENFTLDPAD

51 FFLNCCDCCS PVF **GQKGEPG ETGQPGPKGE AGNLGIPGPP GVVGPQGPRG**
Collagen region

101 **YKGEKGEPGP KGDKGNIGLG GVKGQKSGK DTCGNCTKGE KGDQGAMGSP**
Collagen region

151 **GLHGGPGAAG EKGEMGEKGE MGDKGCCGDS GERGGKGQKG EGGMKGEKGS**
Collagen region

201 **KGDSGMEGKS GRNGLPGAAG DPGIKGEKGE LGPPGLLGPT GPKGDIGNKG**

251 **VRGPTGKKGS RGFKG**

C)

1 SKGELARVPR **SAFSA GLSKP FPPNIPK EKI LYNDQGN YSPVTGKENC**
C1q domain

51 **SIPGT YVFSY HITVRGRPAR ISLVAONKKQ FKSRETLYGO EIDQASLVI**
C1q domain

101 **LKLSAGDQVW LEVSKDWNGV YVSAEDDSIF TGFLI** YPEET SGISP

7/10

Figure 6

hg15_dna	1697	GCATATCATTCACTTCTTATGATTCAATACATATATATTTCTTTTATATG	1746
insp161.cdna	1	GCATATCATTCACTTCTTATGATTCAATACATATATATTTCTTTTATATG	50
hg15_dna	1747	AGTATATGTTTATATGACTTAGTCTTTTCATTGGACCTCACATAATATG	1796
insp161.cdna	51	AGTATATGTTTATATGACTTAGTCTTTTCATTGGACCTCACATAATATG	100
INSP161-AP1			
hg15_dna	1797	TATATATTTTCCTATTATATCTTTCTCCAGCTTCAAATATGTGGATGTT	1846
insp161.cdna	101	TATATATTTTCCTATTATATCTTTCTCCAGCTTCAAATATGTGGATGTT	150
hg15_dna	1847	TTCTTGGCTTTGTGCTATTTTAATTATTTTGGCTATTGCTGGTATGAACA	1896
insp161.cdna	151	TTCTTGGCTTTGTGCTATTTTAATTATTTTGGCTATTGCTGGTATGAACA	200
hg15_dna	1897	CAATAGCAAAGACCACACCACATACCAAATTTACGAAGAAATCTGAGGAA	1946
insp161.cdna	201	CAATAGCAAAGACCACACCACATACCAAATTTACGAAGAAATCTGAGGAA	250
hg15_dna	1947	AGAGAGATGCCAAAGGGTCTAAAGCCATCCAGTGGCCACCTCCAGAAGA	1996
insp161.cdna	251	AGAGAGATGCCAAAGGGTCTAAAGCCATCCAGTGGCCACCTCCAGAAGA	300
hg15_dna	1997	AGAAGAAACCTCTTCACAGAAATGGCTGAAATGGCAGAACCAATTACCA	2046
insp161.cdna	301	AGAAGAAACCTCTTCACAGAAATGGCTGAAATGGCAGAACCAATTACCA	350
hg15_dna	2047	AACCCCTCGGCCTTGGATTCTGTCTTTGGCACTGCCACTCTCTCTCCCTTT	2096
insp161.cdna	351	AACCCCTCGGCCTTGGATTCTGTCTTTGGCACTGCCACTCTCTCTCCCTTT	400
hg15_dna	2097	GAAAACCTCACTCTTGACCCAGCTGATTTCTTTTGAATTGTTGTGATTG	2146
insp161.cdna	401	GAAAACCTCACTCTTGACCCAGCTGATTTCTTTTGAATTGTTGTGATTG	450
hg15_dna	2147	TTGTTACCTGTACCCGGGCAGAAAGGAGAACCTGGAGAGACTGGACAGC	2196
insp161.cdna	451	TTGTTACCTGTACCCGGGCAGAAAGGAGAACCTGGAGAGACTGGACAGC	500
INSP161-AP2			
INSP161-AP3			
hg15_dna	2197	CAGgtatt.....tctagGTCCTAAAGGAGAGGCTGGAAATTTGGGGAT	4229
insp161.cdna	501	CAG.....GTCCTAAAGGAGAGGCTGGAAATTTGGGGAT	534
hg15_dna	4230	CCCAGGGCCACCAGGAGTTGTTGGGCCCCAAGGCCCTAGAGGCTACAAAG	4279
insp161.cdna	535	CCCAGGGCCACCAGGAGTTGTTGGGCCCCAAGGCCCTAGAGGCTACAAAG	584
INSP161-AP4			
INSP161-AP5			
hg15_dna	4280	GAGAGAAAGgtagg.....ttcagGTGAACCTGGCCCTAAGGGAGATAA	8078
insp161.cdna	585	GAGAGAAAG.....GTGAACCTGGCCCTAAGGGAGATAA	618

8/10

hg15_dna	8079	AGGAAACATTGGTTTGGGAGGAGTGAAAGGACAAAAGGCTCCAAGGGAG	8128
insp161.cdna	619	AGGAAACATTGGTTTGGGAGGAGTGAAAGGACAAAAGGCTCCAAGGGAG	668
hg15_dna	8129	ACACATGTGGGAATTGTACCAAAGGAGAAAAAGGAGACCAAGGGGCTATG	8178
insp161.cdna	669	ACACATGTGGGAATTGTACCAAAGGAGAAAAAGGAGACCAAGGGGCTATG	718
hg15_dna	8179	GGCTCACCTGGCCTGCACGGAGGGCCTGGCGCCAAGGGAGAGAAGGGGGA	8228
insp161.cdna	719	GGCTCACCTGGCCTGCACGGAGGGCCTGGCGCCAAGGGAGAGAAGGGGGA	768
hg15_dna	8229	GATGGGGGAGAAGGGGGAGATGGGGGATAAGGGCTGCTGTGGAGATTCTG	8278
insp161.cdna	769	GATGGGGGAGAAGGGGGAGATGGGGGATAAGGGCTGCTGTGGAGATTCTG	818
hg15_dna	8279	GGGAGAGGGGAGGAAAAGGACAGAAAGGTGAGGGGGGTATGAAAGGGGAA	8328
insp161.cdna	819	GGGAGAGGGGAGGAAAAGGACAGAAAGGTGAGGGGGGTATGAAAGGGGAA	868
hg15_dna	8329	AAAGGTAGCAAAGGAGACAGTGAATGGAAGGCAAAAGCGCCGTAATGG	8378
insp161.cdna	869	AAAGGTAGCAAAGGAGACAGTGAATGGAAGGCAAAAGCGCCGTAATGG	918
hg15_dna	8379	TCTGCCTGGGGCCAAAGGTGATCCAGGGATTAAAGGAGAAAAAGGAGAGT	8428
insp161.cdna	919	TCTGCCTGGGGCCAAAGGTGATCCAGGGATTAAAGGAGAAAAAGGAGAGT	968
hg15_dna	8429	TAGGTCCTCCTGGTCTCCTGGGACCTACTGGGCCGAAGGGTGACATTGGC	8478
insp161.cdna	969	TAGGTCCTCCTGGTCTCCTGGGACCTACTGGGCCGAAGGGTGACATTGGC	1018
hg15_dna	8479	AACAAAGGGGTCCGAGGCCCACTGGGAAGAAGGGCTCTCGGGGCTTTAA	8528
insp161.cdna	1019	AACAAAGGGGTCCGAGGCCCACTGGGAAGAAGGGCTCTCGGGGCTTTAA	1068
hg15_dna	8529	AGGCTCCAAGGGTGAGTTGGCTAGAGTGCCCCGGTCGGCTTTCAGCGCTG	8578
insp161.cdna	1069	AGGCTCCAAGGGTGAGTTGGCTAGAGTGCCCCGGTCGGCTTTCAGCGCTG	1118
hg15_dna	8579	GTTTGTCAAAGCCATTTCTCCTCCTAACATCCCCATCAAATTTGAAAAG	8628
insp161.cdna	1119	GTTTGTCAAAGCCATTTCTCCTCCTAACATCCCCATCAAATTTGAAAAG	1168
hg15_dna	8629	ATTCTCTATAATGACCAAGGGAATTACAGTCCTGTCACTGGGAAGTTTAA	8678
insp161.cdna	1169	ATTCTCTATAATGACCAAGGGAATTACAGTCCTGTCACTGGGAAGTTTAA	1218
hg15_dna	8679	CTGCTCTATTCTGGGACATATGTTTTTCTACCATATTACGGTGAGGG	8728
insp161.cdna	1219	CTGCTCTATTCTGGGACATATGTTTTTCTACCATATTACGGTGAGGG	1268
hg15_dna	8729	GGCGACCTGCTCGAATCAGTCTGGTGGCCCAGAATAAGAAGCAGTTCAAG	8778
insp161.cdna	1269	GGCGACCTGCTCGAATCAGTCTGGTGGCCCAGAATAAGAAGCAGTTCAAG	1318
hg15_dna	8779	TCCAGAGAACTCTCTATGGTCAGGAAATAGACCAGGCCTCTCTCCTCGT	8828
insp161.cdna	1319	TCCAGAGAACTCTCTATGGTCAGGAAATAGACCAGGCCTCTCTCCTCGT	1368
hg15_dna	8829	CATCTTGAAATTAAGTGCAGGAGACCAAGTCTGGCTTGAGGTGTCAAAG	8878
insp161.cdna	1369	CATCTTGAAATTAAGTGCAGGAGACCAAGTCTGGCTTGAGGTGTCAAAG	1418

9/10

hg15_dna	8879	ATTGGAATGGGGTGTATGTCAGTGCTGAGGATGACAGCATT	8928
insp161.cdna	1419	ATTGGAATGGGGTGTATGTCAGTGCTGAGGATGACAGCATT	1468
hg15_dna	8929	TTCCTTTTGTACCCAGAGGAACTTCTGGAATTCACCATAAAT	8978
insp161.cdna	1469	TTCCTTTTGTACCCAGAGGAACTTCTGGAATTCACCATAAAT	1518
		←	
		INSP161-AP6	
hg15_dna	8979	CCTGAATCCTGTAGTTTAGATTTCAGTGAATAAGTCAGTTAACACAGAGT	9028
insp161.cdna	1519	CCTGAATCCTGTAGTTTAGATTTCAGTGAATAAGTCAGTTAACACAGAGT	1568
hg15_dna	9029	AGTGCTATTAAAAAATAAATTCCATTTTCAAGATTAT	9067
insp161.cdna	1569	AGTGCTATTAAAAAATAAATTCCATTTTCAAGATTAT	1607

Figure 7

